DOCUMENT RESUME

ED 104 926 TM 004 382

AUTHOR Cooley, William W.
TITLE Evaluation Methodology for School Practices.
PUB DATE Sep 74
NOTE 19p.; Paper presented at the Annual Convention of the American Psychological Association (82nd, New Orleans, La., Sept. 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Academic Achievement; *Evaluation Methods; Instruction; Intellectual Development; Literature Reviews; Longitudinal Studies; Predictive Measurement; *Program Evaluation; Research Needs; Schools; *Success Factors; Values

ABSTRACT

This paper deals with two fundamental problems that are embedded in the evaluation of school practices: (1) attributing value to outcome measures, and (2) attributing outcomes to particular practices. Ways of generating evidence regarding the value of an outcome are discussed, particularly longitudinal research that clarifies the causal relations that exist between an educational outcome and subsequent events. As far as attributing outcomes to practices is concerned, it is recommended that the instructional practices under investigation be considered as domains of observable dimensions that can explain variation in student outcomes, rather than as a set of discrete, unique practices. A model is called for that can guide the further development of such dimensions. (Author)

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE-
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

ED104926

Evaluation Methodology for School Practices

William W. Cooley

I have been asked to consider with you today some of the broad issues in evaluation methodology, particularly as they relate to needed research and development in educational psychology. Specifically, I want to deal with two fundamental problems that are embedded in the evaluation of school practices:

1. Attributing value to outcome measures.
2. Attributing outcome effects to particular school practices.

Attributing Value to Outcomes

First, I shall deal with the problem of establishing a valued set of outcome measures. In the past, innovators have been rather inarticulate regarding the value of the outcomes of their new curricula or programs. Where clear objectives did exist, evaluators did not question where they came from nor assessed their value. The task of evaluation was thought to include only the following steps:

Invited address presented at the meeting of the American Psychological Association, New Orleans, September 1974.

TM 004 382

review the stated objectives of the educational program, locate or develop measures of those objectives, and then determine how well the program achieved those objectives. In this approach, the main question asked about the outcome measure was how well it assessed the objectives of the program. Clearly we must expand this approach by recognizing that value statements regarding school outcomes are also subject to empirical investigation. If real progress is to be made in evaluating school programs, it is important for us to "demystify" the value field, since all planned educational practice is influenced by the value attached to the educational ends that are sought.

Dewey's (1939) Theory of Valuation is a most convincing guide into value inquiry. His central argument is that facts have a bearing on values. Dewey put it this way, "The notion that valuations do not exist in empirical fact, and that therefore value conceptions have to be imported from a source outside experience, is one of the most curious beliefs the mind of man has ever entertained [p. 58]." To assert that value propositions are somehow different from other kinds of propositions and are not empirically testable is to be at the mercy of special interest groups who commend values favorable to themselves as universals. Value disputes are then settled by power alone, rather than by rational inquiry. Dewey

wanted us to see that general propositions regarding value can and should be grounded in empirically tested cause-and-effect relationships.

Means-ends relations are cause-effect relations, and they form the same type of continuum, with means leading to ends which, in turn, are means to subsequent ends in a seemingly endless chain. A particular outcome measure will be highly valued if it is perceived as a necessary link in a means-end continuum leading to some end-in-view, which arises from a currently perceived need. The end-in-view, according to Dewey, is what is desired to meet that need, and is continually revised based on the consequences of our experience. The end actually attained becomes the means for achieving future ends.

There have been serious attempts by psychologists to treat value and valuation objectively. Hull, in his 1952 book entitled A Behavior System, devoted an entire chapter to "Value, Valuation and Behavior Theory." He showed how Dewey's general approach could be the basis for a psychological theory of values and valuing. Another example is in Plans and the Structure of Behavior, where Miller, Galanter, and Pribram (1960) presented a model of human behavior that deals directly with values. They proposed that each of us establishes an internal cause-and-effect representation or

model of the universe. This cognitive map of how the world works, which they called the Image, was thought to control our actions. Miller and his colleagues further proposed that values are part of this Image, and are needed in order to cope with the problem of choice. They pointed out that values are based upon the empirical knowledge that forms the person's Image.

Individuals do operate as if they had a causal model of the world that guides their behavior. One contribution of attribution theory is to reveal the different kinds of causal models that people have. When people have different values, it is because they have built up different models. Some individuals, for example, believe that in our present society, schooling is the great certification mechanism for better jobs. They also believe that success in school depends upon the development of general academic ability. Therefore, they value school programs that are shown to positively affect the development of academic ability. They are not impressed with claims that such abilities are not related to success as adults, because their causal model tells them that such abilities are very relevant.

Similarly, some people believe that differences among schools, teachers, educational programs, etc., do make a difference in what children learn. Thus, they are not particularly impressed when they

are told that school differences do not affect achievement. Few parents seem to be willing to run the risk of having their children attend "poor schools" solely on the basis of one study indicating that school differences have no effect on the development of academic ability. They value good schools because they believe good schools lead to better abilities, which lead to better jobs, more meaningful lives, etc.

In order to illustrate the kinds of research that I am calling for here, let us examine the evidence regarding the value of general intellectual development as a school outcome. General intellectual development is a summary of the current status of the cognitive development of the individual student. In a sense, it is the current profile level for a variety of cognitive tasks that sample different aspects of scholastic ability. It can be measured as the principal component of just about any battery of cognitive tasks currently available. The exact nature of the battery is not important because of the pervasive nature of this general factor, as shown by the high degree to which the principal component from one battery correlates with that from another.

The transfer value of general intellectual development within schooling is extremely well established. No one seems to doubt that the general factor, when measured at one educational level, is by

far the best predictor of academic performance at the next level. Its established within-school predictive validity is the main reason why it is the primary basis for college admissions decisions.

Where the value of general intellectual development seems to be controversial is in its relationship to the extra-school performance of both students and individuals who have completed formal training. Some studies, such as those by Holland and Richards (1965, 1967) and Wallach and Wing (1969), reveal a certain lack of relationship between general intellectual development and selected talented accomplishments that occur outside the classroom. Hoyt (1965) summarized 46 studies that, on the whole, suggested that college grades tend not to be related to earnings within occupational groups. Jencks et al. (1972) reviewed these kinds of studies and extrapolated to the generalization that success in schooling is unrelated to success as an adult. In a report of some recently completed research, Munday and Davis (1974) concluded that, "As we evaluate college outcomes in terms of postcollege student behaviors, we may have to reappraise the central role previously assigned academic talent [p. i]."

In contrast, the Project TALENT research (e.g., Cooley & Lohnes, 1968) showed that general intellectual development is by far the most important single predictor of the occupational sorting that takes place subsequent to the termination of schooling. Although

it may not be highly related to criteria such as earnings within an occupational group, it quite clearly distinguishes among occupational groups. It is even quite highly related to occupational group membership for people who have reached the same level of educational attainment.

These studies illustrate that the relationship between general intellectual ability and adult accomplishment is clearly a researchable question, and is very much in need of further clarification. The kind of research that is relevant to the value of an outcome measure is longitudinal research that clarifies the causal relations that exist between the educational outcome and subsequent events. It is one kind of research that must be undertaken if there is to be a convincing basis for the selection of appropriate dependent variables for studying the effects of school programs. If we want others to accept new outcome measures, their value must be demonstrated.

It is important to note at this point that a measure's value as an indicator of school outcomes must be distinguished from its utility for other purposes. For example:

1. General intellectual development is not necessarily useful for making instructional decisions regarding individual children.

2. It is not necessarily valid for screening applicants for particular kinds of jobs.
3. Other, more specific measures may be better predictors of specific extra-school performances and have better validity.

Certainly, other measures are more appropriate than general academic talent in many situations, but that fact is not relevant to the value of general intellectual development as a general measure of school outcomes. Abuses in the application of a measure represent a poor argument for never using it.

In addition to establishing an outcome's predictive validity in longitudinal studies, there are other ways to generate evidence regarding the value of an outcome measure. For example, an analysis of the kinds of knowledge and abilities needed by adults that are not now being provided by schools would suggest outcomes that would be valued. One could also determine valued outcomes through inquiry into the needs of society, and the current needs of students while they are in school, as opposed to needs following graduation.

In selecting outcome measures, there are at least two other criteria that must be considered in addition to value. First of all, there must be some theoretical or empirical basis for expecting that the outcome being measured can be affected by the school program

or practices under consideration. If, for example, general intellectual development were purely a function of one's biological equipment, then it would not be an appropriate outcome measure for assessing the effects of school programs. But since it has been shown to be, at least in part, a function of environmental conditions, it can be used as an outcome. Secondly, for reasons of interpretive simplicity, outcomes must be limited to a small set of relatively independent measures. Elsewhere, Lohnes and I outline the reasons for this and provide an example of such a set derived from the Project TALENT battery (Cooley & Lohnes, in preparation). It seems to be widely recognized that a broader spectrum of outcomes must be considered in the evaluation of school practices, but this will not happen unless such measures are built and their value demonstrated.

The problem of attributing value to outcomes clearly goes beyond what has traditionally been included in evaluation research, but it seems to me to be the area in greatest need of clarification today. Here I have focused upon just one aspect of the problem, establishing the transfer value of outcomes. As we begin to delve more deeply into this area, other aspects will have to be dealt with, such as the establishment of need, which surely must go beyond the kind of superficial survey that characterizes needs-assessment today. Another concern that must be investigated is the way in which new

facts, which have a bearing upon values, become part of one's cause-effect model of the world. It is one thing to provide evidence relevant to the causal relations between school outcomes and subsequent events, but quite another to have individuals incorporate these relations into their causal model of the world, particularly if the relations are inconsistent with their own model.

Attributing Outcomes to Practices

Given a set of valued outcomes, the problem of attributing outcome effects to particular school practices remains. In the past, the most typical approach to this problem has been to follow the dictates of experimental design. Experimental and control groups were established, with the experimental group participating in the new school practice, while the control group followed the "traditional" one. Whenever possible, randomization was introduced in an attempt to "control" the "uncontrollable." Different curricula or different school practices were treated as if they were quite different and distinct. Thus, if brand A was compared to brand B, the ways in which A and B were similar or different went unnoticed, as did the ways in which schools and teachers differed in the manner in which they used either A or B.

One of the most serious difficulties with this general approach is that new school practices established in a variety of classrooms

by a variety of teachers cannot be represented as a homogeneous treatment in a fixed effects design. It is very safe to assume that new educational practices will vary in significant ways when implemented in two or more classrooms. In view of this known variation in implementation, a major challenge that faces evaluative research is to devise good measures of program implementation, and to develop an acceptable methodology for sorting out treatment effects when treatments vary in uncontrollable ways.

The recommendation developed here is that the instructional practices under investigation in a particular program of evaluative research be considered as a multidimensional domain, rather than as a set of discrete, unique practices. That is, different curricula and instructional methods should be conceived of as differing in degrees along common dimensions, rather than differing in kind or type.

There have been many important attempts to define observable dimensions of instructional practices. Research conducted by the Stanford Research Institute (SRI) on the Follow Through program (e.g., Stallings et al., 1973) is one current example. Their classroom observation scheme clearly reveals the differences among the various instructional models participating in Follow Through. Leinhardt's (1974) research at the Learning Research and Development Center is defining variables that are relevant to variation in

the implementation of a particular instructional model. Her work is based on assessments of the key features of that model, together with guesses as to the kinds of distortions in the model one was likely to find in the field that would affect the model's effectiveness. Rosenshine (1971) provides an excellent summary of variables that can be derived from teacher-effectiveness research, which has been going on for decades. This research has yielded an assortment of variables ranging from teacher warmth to the businesslike character of the classroom, with some beginnings toward consistency among the trends.

What is needed now is a theory to guide the further development of instructional dimensions. If evaluative inquiry is so designed, it could have very important implications for a theory of instruction as well as for educational policy. Of course, the purpose of evaluative inquiry is to provide policy implications, but it could also have theoretical implications if it is guided by and designed in terms of available theory, no matter how primitive it might be. Also, evaluative inquiry might be more convincing as a basis for policy if it is considered in a theoretical framework.

A problem, of course, is the fact that attempts to develop theories of instruction are a relatively recent phenomenon, and theories that are presently available are indeed primitive. Bruner

(1966) has proposed that a theory of instruction will be quite different from available theories of learning. He suggested that instructional theory is a prescriptive theory for the optimization of learning, as opposed to a descriptive theory of learning that has occurred. Glaser's (1970) model for designing and building an instructional program is an excellent example of that prescriptive approach.

In contrast, Carroll (1963) has proposed a conceptual model of the factors that affect success in school learning. The basic proposition in Carroll's model is that "the learner will suceed in learning a given task to the extent that he spends the amount of time that he needs to learn the task." Essentially, his quite simple model proposed that the degree of learning that takes place in the course of instruction is a function of the time actually spent in direct learning activities divided by the time needed. Further, he proposed that the time actually spent in learning would be a function of the opportunity provided for learning, the amount of time the learner was willing to actively engage in learning, and the amount of time that was needed to learn, plus an increment that was a function of the quality of instruction or the student's lack of ability to understand less than optimal instruction.

In dealing with the problem of attributing outcome effects to particular school practices, a Carroll-type approach will probably

have greater utility than a prescriptive model. What is sought is a model of the instructional domain that will specify the kind and amount of learning that will occur in a given educational environment, typically a classroom. In a reanalysis of some of the SRI Follow Through data, Emrick and I have shown that a modification of the Carroll model is very useful in organizing a large number of classroom variables into a manageable set of four classroom descriptors which, in turn, are very useful explanatory measures of outcome effects (Cooley & Emrick, 1974). The four classroom dimensions were based upon available SRI variables thought to be related to opportunity to learn, degree of curriculum structure, quality of the instructional events, and classroom events increasing student motivation. What remains to be done is to derive a set of classroom measures from such a conceptual model (as opposed to selecting a set from an available data bank) and to further test its validity using a variety of classroom settings.

The purpose of this discussion of classroom dimensions has been to illustrate the kind of model that is essential if we are going to move forward our ability to attribute outcome effects to educational practices. A convincing theory or conceptual model of the phenomena is much more essential to causal inference than is, for example, random assignment of subjects to treatments in an orthogonal design. As indicated earlier, in studying school practices

the design does not stay orthogonal anyway because of implementation variation. Also, laboratory-like controls may introduce artificialities that limit generalizability to field situations. This problem is far more serious than having certain ambiguities introduced through less than perfect experimental control, particularly since we can guard against generalizing about non-orthogonal confounding if we directly measure the treatment dimensions. Of course, there are those who argue that our knowledge of what is important in classroom environments is still too primitive to allow us to construct a convincing model. I believe that the development of a model is now possible and desirable. It is certainly essential that we reduce the hundreds of variables that investigators have considered in their attempts to find ways in which classroom/teacher differences make a difference in what children learn. Doing so will allow us to make significant progress in the evaluation of school programs.

But Is It Evaluation?

I suspect that for many of you I have strayed way beyond a consideration of evaluation methodology. What I have done is suggest that some rather fundamental research is required before we will significantly improve the evaluation of school practices. I don't mean to assign such research to people called "evaluator," nor necessarily to include it in a set of activities called evaluation research. Research

on the transfer value of outcomes and the development of a model of classroom learning that can explain variation in student outcomes represent a set of concerns that should be very relevant to psychologists interested in applying their discipline to critical educational problems.

References

Bruner, J. Toward a theory of instruction. Cambridge, Mass.: Belnap Press, 1966.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Cooley, W. W., & Emrick, J. A. A model of classroom differences which explains variation in classroom achievement. Paper presented at the meeting of the American Educational Research Association, Chicago, April 1974.

Cooley, W. W., & Lohnes, P. R. Predicting development of young adults. Palo Alto, Calif.: Project TALENT, 1968.

Cooley, W. W., & Lohnes, P. R. Evaluative inquiry in education, in preparation.

Dewey, J. Theory of valuation. Chicago: University of Chicago Press, 1939. (International Encyclopedia of Unified Science, Volume II, Number 4)

Glaser, R. Evaluation of instruction and changing educational models. In M. C. Wittrock and D. E. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart and Winston, 1970. Pp. 70-86.

Holland, J. L., & Richards, J. M., Jr. Academic and nonacademic accomplishment: Correlated or uncorrelated. Iowa City, Iowa: American College Testing Program, 1965. (ACT Research Report No. 2)

Holland, J. L., & Richards, J. M., Jr. Academic and nonacademic accomplishments in a representative sample of students taking the American College Tests. College and University, 1967, 43, 60-71.

Hoyt, D. P. The relationship between college grades and adult achievement: A review of the literature. Iowa City, Iowa: American College Testing Program, 1965. (ACT Research Report No. 7)

Hull, C. L. A behavior system: An introduction to behavior theory concerning the individual organism. New Haven, Conn.: Yale University Press, 1952.

Jencks, C., et al. Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

Leinhardt, G. Evaluation of the implementation of a program of adaptive education at the second grade (1972-73). Paper presented at the meeting of the American Educational Research Association, Chicago, April 1974.

Miller, G. A., Galanter, E., & Pribram, K. H. Plans and the structure of behavior. New York: Henry Holt and Company, 1960.

Munday, L. A., & Davis, J. C. Varieties of accomplishment after college: Perspectives on the meaning of academic talent. Iowa City, Iowa: American College Testing Program, 1974. (ACT Research Report No. 62)

Rosenshine, B. Teaching behaviours and student achievement. London: National Foundation for Educational Research in England and Wales, 1971.

Stallings, J. A., et al. Follow Through program classroom observation evaluation 1971-72. Menlo Park, Calif.: Stanford Research Institute, 1973.

Wallach, M. A., & Wing, C. W., Jr. The talented student. New York: Holt, Rinehart and Winston, 1969.